Expectation-Maximization Algorithm

February 21, 2019

1 INTRODUCTION TO EXPECTATION-MAXIMIZATION(EM) ALGORITHM

1.1 An example of EM algorithm

Let's start with an example from Chuong B Do & Serafim Batzoglou's tutorial paper *What is the expectation maximization algorithm?*, and I have saved the .pdf file of this paper in the *material* folder.

If we can figure out whether the coin we flipped is A or B, we can estimate θ_A and θ_B through one of our most familiar friends, **MLE**.(Figure 1.1a) But actually, since we can not figure it out, MLE will not work in this situation. Then EM algorithm come on the stage. EM algorithm is simple because it just contain two steps in the Iteration and EM algorithm is complex is complex for its mathematical proof and further stretch in theory and application (Figure 1.1b) andI will explain the two steps in the following section.

1.2 Two steps in iteration of EM

Given the statistical model which generates a set X of observed data(5 samples in coin-flipping experiment, and x_i = # of heads using the 1^{st} coin), a set of unobserved latent data or missing values Z(the coin is A or B), and a vector of unknown parameters θ, along with a likelihood function $L(\theta; X, Z) = p(X, Z|\theta)$, and the marginal likelihood function of the observed data is given by:

$$L(\theta; X) = p(X|\theta) = \int p(X, Z|\theta) dZ$$

However, this quantity is often intractable (we don't know the coin we tossed is A or B).

Then the EM algorithm seeks to find the MLE of the marginal likelihood by iteratively applying the following two steps.

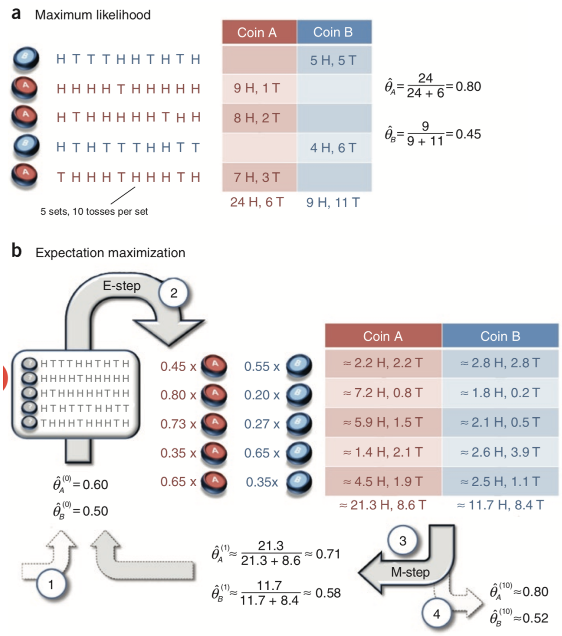

Figure 1.1: Example: As an example, consider a simple coin-flipping experiment in which we are given a pair of coins A and B of unknown biases, θ_A and θ_B, respectively. Our goal is to estimate $\theta = (\theta_A, \theta_B)$ by repeating the following procedure five times: randomly choose one of the two coins (with equal probability), and perform ten independent coin tosses with the selected coin. Thus, the entire procedure involves a total of 50 coin tosses'

(a) Expectation Step: Define $Q(\theta|\theta^{(t)}$ as the expected value of the log likelihood function of θ, with respect to the current conditional distribution of Z given X and the current estimates of the parametrers $\theta^{(t)}$:

$$Q(\theta|\theta^{(t)}) = E_{Z|X,\theta^{(t)}}[logL(\theta; X, Z)] = \int P(Z|X, \theta^{(t)}) lnP(X, Z|\theta) dz$$

note: In discrete situation, we replace $\int$ by $\sum$. In the coin-flipping experiment, we assume $\theta^{(0)} = (\theta_A^{(0)}, \theta_B^{(0)}) = (0.60, 0.50)$. Since the results of the first ten independent tosses is $x_1 = 5$ ({H: 5, T: 5}), we have $P(Z = A, X = x_1, \theta^{(0)}) = 0.6^5 \cdot 0.4^5$ and $P(Z = B, X = x_1, \theta^{(0)}) = 0.5^{10}$. Further,

$$P(Z = A|X = x_1, \theta^{(0)}) = \frac{P(Z = A, X = 5, \theta^{(0)})}{P(Z = A, X = 5, \theta^{(0)}) + P(Z = A, X = 5, \theta^{(0)})} \approx 0.45,$$

$$P(Z = B|X = x_1, \theta^{(0)}) = \frac{P(Z = B, X = 5, \theta^{(0)})}{P(Z = B, X = 5, \theta^{(0)}) + P(Z = A, X = 5, \theta^{(0)})} \approx 0.55$$